DOCUMENT RESUME

ED 449 771 HE 033 783

AUTHOR	Mentkowski, Marcia; Astin, Alexander W.; Ewell, Peter T.; Moran, E. Thomas
TITLE	Catching Theory Up with Practice: Conceptual Frameworks for Assessment. The AAHE Assessment Forum.
INSTITUTION	American Association for Higher Education, Washington, DC.
PUB DATE	1991-00-00
NOTE	41p.; Foreword by K. Patricia Cross. Based on a session from the 1991 Association for the Study of Higher Education Assessment Conference, subsequently enriched and updated.
AVAILABLE FROM	American Association for the Study of Higher Education, One Dupont Circle, Suite 600, Washington, DC 20036-1110 ($10). Tel: 202-293-6440; Fax: 202-293-0073.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Educational Assessment; Educational Philosophy; Higher Education; *Student Evaluation; *Theory Practice Relationship
IDENTIFIERS	*Reflective Thinking

ABSTRACT

The four participants in this "conversation" make the case for better theory to support contemporary practice in assessment. The participants tell individual stories about problems in assessment and then explore the nature of measurement. As a whole, the conversation embodies reflection as a central value of assessment. The assessment movement used to focus on an intellectual and functional coherence among educational goals, curricula, and outcomes, but this conversation shows that it is now important to explore the need for coherence between assessment theory and practice. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 449 771

THE AAHE ASSESSMENT FORUM

AMERICAN ASSOCIATION FOR HIGHER EDUCATION

CATCHING THEORY UP WITH PRACTICE: CONCEPTUAL FRAMEWORKS FOR ASSESSMENT

by Marcia Mentkowski, Alexander W. Astin,
Peter T. Ewell, and E. Thomas Moran

with a foreword by K. Patricia Cross

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B. Pollack

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

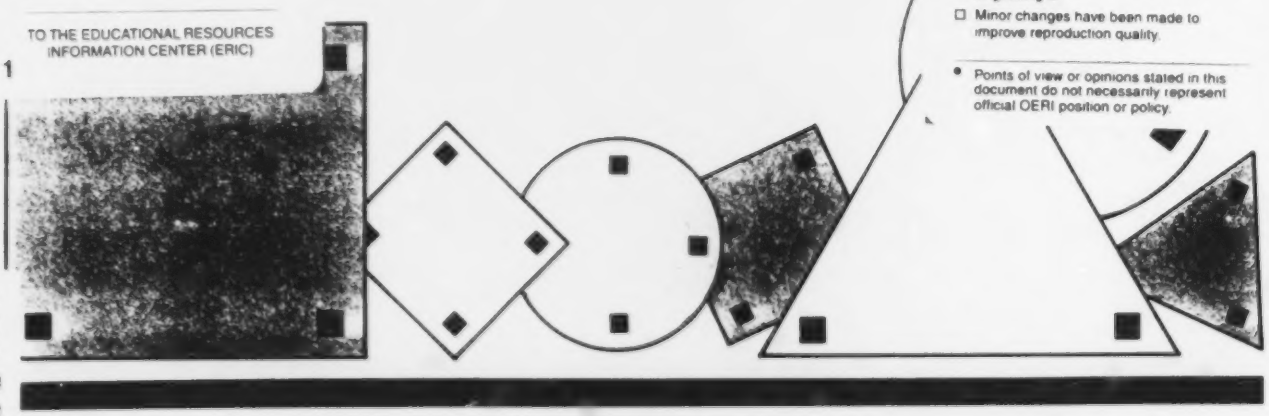
1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



HE033783

BEST COPY AVAILABLE

CATCHING THEORY UP WITH PRACTICE: CONCEPTUAL FRAMEWORKS FOR ASSESSMENT

by Marcia Mentkowski, Alexander W. Astin,
Peter T. Ewell, and E. Thomas Moran

with a foreword by K. Patricia Cross

The AAHE Assessment Forum
American Association for Higher Education
1991

For more about the American Association for Higher Education and its Assessment Forum, see pages 30-33.

This publication may not be reproduced without the permission of the American Association for Higher Education, One Dupont Circle, Suite 360, Washington, DC 20036-1110; ph. 202/293-6440, fax 202/293-0073.

Additional copies of this publication are available for $10 each, including 4th Class postage. Orders under $50 must be prepaid; orders over $50 must be accompanied by payment or institutional purchase order. Allow 6 to 8 weeks for delivery. Order from:
AMERICAN ASSOCIATION FOR HIGHER EDUCATION
One Dupont Circle, Suite 360
Washington, DC 20036-1110
ph. 202/293-6440, fax 202/293-0073

Foreword

The 1980s were a difficult decade for education. In 1986, in the midst of widespread criticism of all levels of education, the nation's governors in *Time for Results* charged that "today's graduates are not as well educated as students of past decades." Observing that "not enough is known about the skills and knowledge of the average college graduate," the governors pressed hard for continuation and strengthening of the emerging assessment movement.

The 1990s promise to be a decade in which higher education takes charge of its own agenda. More than anyone else, educators want to know how effective colleges are in educating students, and many educators are deeply committed to improving the quality of undergraduate education. Although academics are often accused of being long on talk and short on action, the nationwide action-oriented assessment movement belies that charge. Higher education has moved with extraordinary speed, given the complexity of the task and the number of people involved, to put assessment programs into action. By 1990, 82 percent of the colleges and universities surveyed by the American Council on Education had some form of assessment activity under way, and two thirds were developing their own instruments for student assessment. While these efforts may not yet have penetrated very deeply into everyday academic life, still, they are impressive beginnings.

While it is perhaps not quite fair to say that more people are *doing* assessment today than are *thinking* conceptually about it, there is little doubt that practice is moving faster than theory. As practitioners forge ahead to develop assessment programs, they keep bumping up against deep and troubling questions that can be addressed only by developing stronger conceptual frameworks. Before those conducting the assessment can decide how to *do* assessment, they should at least try to agree on what they *think* about education. "We need to put more emphasis," says Marcia Mentkowski in these pages, "on assessment as part of the educational process and ask ourselves what, exactly, is driving assessment. Is it just an administrative priority that we can stop worrying about because someone else is taking care of it? Or is assessment being driven by the educational process, [by the] assumptions, values, and questions of the faculty?"

This AAHE Assessment Forum publication addresses that growing need to examine our operative assumptions, about education as well as about assessment, and develop appropriate new conceptual frameworks. It has its roots in AAHE's 1989 Assessment Conference, at which Marcia Mentkowski gave a major address on the topic "Catching Theory Up With Practice: Establishing the Validity and Integrity of Higher Education Assessment." Discussion groups that met following her talk found theory an engaging topic for

conversation but had great difficulty articulating the issues that confronted practitioners and scholars alike.

The challenge of catching theory up with practice was next taken up by a quartet of leaders in the assessment movement — Alexander Astin, Peter Ewell, Marcia Mentkowski, and Thomas Moran. These four people come at assessment from very different academic backgrounds, but all are deeply involved with both the practical and the theoretical aspects of assessment. In informal conversations held over a three-year period — some transcribed and analyzed — they have articulated issues and confronted the lack of theoretical underpinnings to support and direct assessment efforts.

Finally, there is a third source for the conversation presented in these pages: a standing-room-only session at AAHE's 1991 Assessment Conference, in San Francisco, organized by Marcia, that dealt with these issues. Barbara Wright, director of the AAHE Assessment Forum, edited the transcript of the session, and that transcript became the skeleton for this piece. Then Marcia, together with Peter, Tom, and Sandy, fleshed out the text, sharpening the issues and enriching the transcript with excerpts from prior discussions. Along the way, all four contributors continuously updated the text, as they carried on the dialogue among themselves and with their colleagues.

The case for better theory is compellingly presented in the opening pages of the dialogue, as the authors tell their individual stories about their own encounters with problems in assessment. Readers will not find in these pages much information about how to *do* assessment; this is a how-to-*think*-about-it conversation. As Sandy Astin notes, "we shouldn't even consider the questions of instrumentation and methodology until we've at least *tried* to answer [the larger questions of] Why are we doing this? What do we hope to learn? How might we use the resulting information? How will we make sense out of it? And how can we get the larger academic community to take an interest in it, ascertain meaning from it, and use it to improve student learning and development?"

But once the larger questions of purpose are addressed, assessors are plunged into the thorny thicket of assumptions about the nature of measurement. Peter Ewell provides numerous examples of traditional assumptions that don't seem to work anymore, but one that will strike a responsive chord with many educators concerns the nature of critical thinking — a priority item on most assessment agendas these days. We have assumed, Peter says, "there was such a thing as critical thinking out there that you had some of, then more of, and we conceived of every individual as having it in more or less the same way. If we conceived of it that way, central-tendency measures made sense. . . . The point here is that *how* you model the ability is an active choice that you have to make." As a matter of fact, notes Sandy Astin, "some of the most interesting change in a cohort of students is in the *variability* rather than the mean. This appears to be happening in such diverse domains as quantitative skills and political beliefs, where students become more *diverse* over time."

The more we think about the methods and uses of assessment, the more the traditional paradigms of educational research come into question. Tom Moran is a critic of the pos-

itivism that underlies the psychometric theory that has driven so much of assessment. "When we look at our students through this methodological lens," he says, "we are not . . . looking at the student as a self-willed actor." As a critic of positivistic, linear models, Tom has plenty of company these days in the revolution in research methods that is raging through the disciplines.

The conversation as a whole embodies a central value of assessment: reflection. Early on, the movement struggled to become more conscious of and explicit about educational goals. Now these four practitioner-theorists elaborate the theoretical underpinnings of the assessment movement itself. Early on, the movement sought to establish an intellectual as well as functional coherence among educational goals, curricula, and outcomes; now they explore the equally fundamental and urgent need for coherence between assessment theory and practice.

The Four Participants

Because the four participants in this conversation come from different disciplines and offer different perspectives, it is useful to know something of the experience that shapes their perceptions.

Marcia Mentkowski is a professor of psychology and director of the Office of Research and Evaluation at Alverno College. She approaches assessment from the perspective of a developmental and educational psychologist concerned about the growth and development of students at a student-centered, outcomes-oriented college. Alverno pioneered in the assessment movement and has been especially successful in making assessment an integral part of teaching and learning. Alverno faculty involve students, alumnae, and community professionals in a continuous assessment process that centers on the individual student and her educational experiences in the classroom, and extends across every level: from students to curriculum to the college as a whole.

Alexander Astin is a psychologist by discipline and a professor of education at the University of California, Los Angeles. Author of *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*, Sandy is committed to moving from a definition of academic quality based on resources or reputation to one based on student achievement or talent development. Sandy is concerned about the values that are embedded in conventional assessment practices; all too often, this means the encouragement of competition over collaboration, elitism over equity, and number crunching over measures of genuine educational improvement.

Peter Ewell, senior associate at the National Center for Higher Education Management Systems (NCHEMS), comes at assessment from an interinstitutional perspective, seeing issues and problems that are common across institutions. He has visited scores of campuses across the educational spectrum and has written a number of articles on the state of the art in the assessment movement. In background and training, Peter is a political scientist with a strong statistical bent specializing in econometrics and survey research.

Thomas Moran is especially interested in assessment as it pertains to institutional mission and purpose. With the support of a FIPSE grant, he has worked extensively to assess learning outcomes in the major field of study across colleges in the SUNY system. Tom has studied administration and policy analysis, with an emphasis on organizational theory, and puts theory into practice as a vice president for academic affairs at the State University of New York at Plattsburgh. He is particularly interested in the paradigm shifts that are taking place in the social sciences, changes that are likely to affect the way we think about assessment in the future.

Readers will find much to pique their interest in this conversation among these exceptionally thoughtful and articulate practitioner-scholars of assessment. On some pages I found insights that shed new light on old questions; elsewhere I found an entirely new set of problems to worry about. I think I can guarantee that virtually anyone involved in the assessment movement — and that should include all educators — will find something to think about in these pages. The primary purpose here is to expand the conversation to engage a wider audience.

K. Patricia Cross
Elizabeth and Edward Conner Professor of Education
University of California, Berkeley

I. UNEASE AND INSIGHT

EWELL: As we have these conversations, we often engage in storytelling. I want to approach our first question — Why do I see a need for new conceptual frameworks for assessment? — by reporting on a particular moment of truth that I experienced. I had this flash of insight at AAHE's Second Conference on Assessment, in Denver in 1987, during a session on the topic of value-added.

There was an enormous debate about value-added. Everybody seemed to be questioning it, but it was not clear what the "it" was that they were rejecting. There were people who objected to value-added on statistical grounds, because of the instability of gain scores, arguing that instead of gain scores, we ought to be using a complex regression model and the residuals from it to estimate how much "learning" took place. There were other people who said, no, that's not right. We object to the whole assumption of linear development that underlies the notion of value-added. We should be looking at patterns of development rather than buying into a statistical model fundamentally rooted in a notion of linear change. But *that* objection still had embedded in it an assumption that learning consisted of abstract, generalizable abilities that might follow established patterns; so a third group of people was objecting on the grounds that what we should really be looking at is *individual* paths of development in specific contexts.

So the whole notion of value-added seemed to be getting exploded from some very different directions. And there were few if any alternatives directly in sight. It seemed to me that we had to completely rethink the question.

ASTIN: I think there's been a lot of clouding of the issues in these discussions of value-added. The basic idea behind value-added (or "talent development," as I prefer to call it) is twofold: that learning and growth takes place over *time* and that assessment cannot hope to document that growth unless it also tries to reflect how students are changing over time. It has very important implications for assessment: It means you can't learn very much from one-time administrations of achievement tests. But simply measuring how much learning takes place over time is not enough; we also need to know *why* some people change more than others. This is why it's so important to develop one's philosophy of assessment. For example, the model of assessment that I develop in my book *Assessment for Excellence* is based on a fundamental assumption: Assessment results are of most value when they shed light on the *causal connections* between educational practice and educational outcomes.

That's why I developed the input-environment-outcome (I-E-O) model. The student's learning or growth is reflected in the change from input (I) to outcome (O), and the environmental information (E) provides an explanatory framework for understanding *why* individual students change as they do. This model is designed to yield assessment results that will simultaneously provide maximum information on the possible causal connections between various educational practices and educational outcomes and minimize the chances that our causal inferences will be wrong.

I've had plenty of occasions over the years to test out this assumption and the model, and it seems to work. But along the way, some of my colleagues who are experts in the methodology of social science or educational research cringed at my frequent references to "causal" relationships or to the "effects" of educational programs. Most of them, like me, were brainwashed during their graduate training about the superiority of "true experiments" over "correlational studies." We were all repeatedly reminded that "you can't make causal inferences from correlational data." It has taken me several decades to realize that all of this well-intended advice is simply wrong: True experiments are no panacea, in part because they are very difficult if not impossible to conduct with live human beings in real educational settings, and in part because they create at least as many inferential problems as they solve. And while it is true that you can't *prove* causation with correlational data, you most certainly *can* make causal inferences from such data; people do it all the time. In fact, it would be impossible for most teachers and administrators to make it through an average work day without making literally dozens of causal inferences based either on correlational data or, as is more often the case, on no data at all. The real challenge for us researchers and practitioners is to use assessment in such a way as to minimize the chances that our causal inferences will be wrong.

MENTKOWSKI: Why do I think it's important to articulate our principles, our philosophies of assessment? For me, it's because I need to know *why* I'm doing assessment the *way* I'm doing it, and what could divert me from my major commitments and values. Otherwise, I may avoid grappling with some of the paradoxes in assessment — paradoxes that threaten to pull me in quite different directions, or worse, dichotomize my thinking without my even realizing it. I think these paradoxes *should* move my thinking forward to new assumptions, rather than catch me up in old polarities.

For example, as an educator and a developmental psychologist, I care very deeply about the kind of teaching, learning, and assessment that students experience. I'm committed to some particular educational assumptions and values. I believe that the outcomes of college should include not only what students know but also what they are able to *do* with their new knowledge. And I value curricula that assist students to expand their definitions of learning: from merely understanding information to a wider definition that includes learning as self-directed, active, collaborative, and experiential. At the same time, and here is where the paradox

comes in, as a researcher and as an assessment professional, I'm responsible to a faculty that trust me to work with them to create multiple ways of examining and evaluating those very educational assumptions and values that we are committed to. We want to know: How do our students learn, how do they *really* turn out, and why?

Now, this has meant that I must work from the outside and the inside, as a member of the institution as well as a mirror of the institution, to both reflect on and to participate in inquiry. And I also need to do a third thing. Together with my colleagues, I must also create and institutionalize a culture of inquiry, a context where reflecting on practice and continuous improvement are part of the way we work. And yet, there are some old assumptions that I learned early on that haunt me.

One of those old assumptions is this: If you have "commitments," you aren't "objective," and your assessment strategies cannot possibly be objective. I learned that in graduate school, and it means that for many years, I've been breaking the rules. Under the old rules, an assessment practitioner is an oxymoron. I needed a new rule, a new assumption, one that does not dichotomize objectivity and subjectivity and that includes a third dimension, as well.

What you have to learn to do — just like in the movie *Annie Hall*, if you are going to be Woody Allen's character — is to stand inside and outside the situation simultaneously. In the movie, you'll recall, Woody Allen plays Annie's boyfriend; he interacts with Annie and simultaneously, for the audience, steps outside of the relationship and evaluates it. And then there's the third part, namely, creating the context for it all to happen and then making it happen: Allen wrote the script and is directing the movie!

Let's stay with that image for a moment. The analogy to my situation is this: I've had to learn to design assessment that could be objective and subjective simultaneously. I've tried to develop what I call a "feelingful" mind, to become interdependent with colleagues in the situation, because I've learned that what we know is interdependent with how we know it. I've learned to step away from the setting, to bring to bear multiple, outside perspectives. And third, I've worked with others to create a context for assessment that works for improvement and accountability — we've written the script and directed the movie with an eye toward the critics, if you will.

But then that old assumption — that if you have a commitment to whomever or whatever is being assessed, you can't be an effective assessor — comes back and bites me. For example, at a conference just a few months ago, Alverno's institutional assessment designs were criticized as an example of traditional empiricism divorced from the concerns of the faculty. The person who said that saw only the "objective" side of the designs. Then later that same week, at another conference, a listener suggested that our institutional assessment designs were an example of assessment totally co-opted by the institution. That person saw only the "subjec-

tive" side of the design. Neither critic recognized the third dimension: the creation of a context where both self-reflective and externally oriented inquiry occur in relation to educational values.

Knowing that we need a new assumption, one that brings objectivity and subjectivity together and builds in that third dimension as well — just knowing that keeps me centered and helps me recognize change when it happens. I've seen this new assumption break through old limitations: instructors taking a systems perspective and evaluating the whole major, administrators building institutional structures that stay centered on the individual learner. In other words, practice suggests new theory, which in turn suggests rethinking practice. Tom?

MORAN: Why do I believe new frameworks are essential? My early training was in social science and I was influenced by the currents of the late 1960s and early 1970s. For example, I was very concerned about reification, about the decoupling of method from substance and technique from meaning.

Nevertheless, when we began working on assessment in the mid-1980s, I approached assessment as if it were "normal" science; in other words, I initially adopted traditional empirical approaches to questions about educational research.

But there were several weaknesses in those approaches that became increasingly clear to me. Chief among them: a unilateral process. We as researchers imposed our own unilateral perception of growth and learning and outcomes upon subjects assumed to have no intentionality — that is, our students.

Here's an example. I began working on what was originally called a "value-added" project. What proved troubling in that project was not its fundamental assumption, namely, that if we care deeply about the impact of college on students, then we have to find some means by which we can determine the contribution of the college curriculum experience to students' development. That seemed sound enough. The real problem was that the conceptual framework underlying value-added got played out so that students were treated as neutral and passive. Significantly, faculty were more sensitive to that problem, at least at the start, than the researchers and administrators who worked on the project, myself included. We failed to appreciate the extent to which students' own effort, the interaction of the institution with them, or the complex interactions within the institution, could alter outcomes for students.

So I began to look at other ways to formulate assessment strategies. What came to mind was the work of William Perry, work I had been familiar with for ten or fifteen years. In *Intellectual and Ethical Development in the College Years,* Perry had developed a framework for assessing cognitive and ethical development among college students. What was fascinating was that he didn't begin by postulating certain outcomes that he expected to find and then testing for them, as a traditional empiricist would. Rather, he engaged in a series of extensive dialogues with students and

allowed their perceptions of the changes that had occurred during their college years to emerge through those dialogues. He was less concerned with cause and effect than he was with the pattern that emerged. The work of William Perry led me to believe that we could develop other models — and more importantly, *frameworks* by which we could justify those other models.

EWELL: I can't help but comment that this brings us around to the value-added observation I started with, Tom. The process that you describe at Plattsburgh was similar to many other projects I saw that called themselves "value-added" projects in the mid- to late-1980s. It was a popular word at the time, and many campuses — Northeast Missouri, Kean College, and others — started with that metaphor but in practice transformed it into something quite different. I guess we were all making up our conceptual frameworks as we went along.

ASTIN: What you're really saying is that using a "value-added" model in a vacuum is not very helpful. Something else is needed; namely, we need to know the *source*: What is it in the student's environmental experience that "adds value"? It was this realization that led me to adopt the input-environment-outcome model in my own work. I use the model in a way that incorporates this basic principle: *assessment involves finding the means to measure the contribution of curriculum and other educational experiences to students.* I find it difficult to see why this idea still gives people problems: that learning and growth takes place over time, and that assessment designs need to document that growth by reflecting how and why students are changing over time.

MENTKOWSKI: It may seem perfectly obvious now, Sandy, but I think this is really a pretty new idea, when we look back over the kinds of college outcomes studies generally done prior to the assessment movement. I took a lot of flak in the mid-1970s when I designed a longitudinal study linking changes in student outcomes to our curriculum, for program assessment purposes. All I did was mirror our faculty's developmental curricular design, but some outsiders were very troubled by it. Take a moment to give us a short history lesson. What's different, and why is it so important to hang on to this idea?

ASTIN: Well, earlier research on the effects of college focused on a lot of things, but not necessarily on the impact of a coherent curriculum on student outcomes. For example, look at the studies summarized in 1969 by Kenneth Feldman and Theodore Newcomb in *The Impact of College on Students*: They tended to compare college-goers with those who didn't go and focused on whether going to college makes a difference. In 1977, in his book *Investment in Learning: The Individual and Social Value of American Higher Education*, Howard Bowen looked at some of this literature and concluded that college did indeed confer important benefits on students. More recently, researchers have gotten interested in identifying some of the very general kinds of college *experiences* that contribute to an overall benefit, such as faculty-student interaction or living on campus. In 1991, Ernest Pascarella and Patrick Terenzini brought us up to date with a review of the last twenty years of this

BEST COPY AVAILABLE

research in *How College Affects Students,* and helped us understand the interplay between students' development and the organizational features of colleges. They provide a broad backdrop against which a college's assessment data can be interpreted. Our most recent research suggests that the *peer group* may be the most important source of influence on students' development.

A significant new wrinkle in all this, supported by the current assessment movement, is to link particular changes in student knowledge, abilities, or values to a particular *curriculum* or educational experience that the faculty as a whole *deliberately* design and implement to enhance student learning. In the past, we were preoccupied with measuring students when they came to college, for admissions or placement purposes. Now we're attuned to measuring outcomes when students graduate. But it would be really naive for us just to ride the pendulum from gathering admissions data to measuring exiting graduates without linking both sets of information to faculty-designed instruction and peer group or cocurricular experiences. And just as naive to overlook the longitudinal part. In other words, we have to make sure we're following the same students so that we have some idea who changed, how they changed, and why. This time around, I'd like to see us put *all* these elements together in our assessment designs, rather than fall back into our old ways of doing things. The benefit is that we will learn a great deal more about how to directly improve a student's educational experience.

II. THEORY, PRACTICE, AND PROBLEMS

MENTKOWSKI: Current assessment practice is actually quite different from what some of us would like to see. It's important for us to identify some of the "syndromes" of current practice. What *are* some of the concrete problems that we face — and how are they linked to a lack of appropriate conceptual frameworks?

EWELL: One problem is that when we talk to external audiences we often use a different language from the one we use when we talk internally, to one another. In fact, we often seem to be talking out of both sides of our mouth, reinforcing the old dichotomies. We talk "quantitative" on one side, "qualitative" on the other; "summative" on one side, "formative" on the other. Instead of breaking up these polarities and clarifying our new assumptions for external audiences, our language and thought cater to the external, and that orientation dominates.

MENTKOWSKI: Right, Peter. Once again we let ourselves get divided between the objective and the subjective. That keeps us looking for outside experts, who are somehow going to be able to come in, rescue us, and solve our problems — maybe psychometricians, or institutional researchers, or external evaluators. Instead, maybe we should think about how to create a culture of inquiry or a community of judgment within our *own* institutions, building in a broad base of expertise.

Now, that doesn't mean that we abdicate responsibility for standard setting, or that we ignore external reference points, such as criteria from a discipline. In fact, it often means *expanding* our community of judgment to include alumni, community professionals, or external examiners. But the idea is to start with our own context, constructing both self-reflective and externally oriented processes, rather than assuming that someone from the outside can solve our problems for us.

MORAN: An exclusively external orientation also produces an excessive focus on the endpoint or outcome of the learning process. This is a failure of much of social science research. It reminds me of the fellow in the Kurt Vonnegut novel who lies on a train looking through a pipe at the Rocky Mountains passing by. He's only seeing a tiny fragment of the grand sweep of the Rocky Mountains.

Our research methods lead us to the same problem in viewing student learning and development: We see only a fragment of the total phenomenon. We don't see learning and development unfolding over time. We don't see continuity or evolution. In fact, any single-point measure, or even multiple data points, leads to the universal problem of cross-sectional data: you know, that great one-liner about how the problem with cross-sectional data is that if you applied it to the city of Miami, you'd conclude that everybody there was born Cuban and grew old Jewish.

Traditionally, we've focused on static, end-point measures alone; we need to include dynamic, longitudinal observations of our students interacting with the curriculum. This means that all of us — faculty, administrators, students — become full participants in the process.

EWELL: I'd like to underscore the point about participation in the process. Although we're often not terribly conscious of it, we face a *choice* here: a choice between buying an available alternative and making up our own instrument. Faced with that choice, we have, for a number of reasons, fallen into excessive reliance on off-the-shelf instruments. Off-the-shelf instruments are easy. Although they cost money, it's a lot less than the effort it takes to develop your own. But buying something off-the-shelf means not really engaging the issues that we should — for example, What are we really assessing? and What assumptions are we making about the nature of learning?

There's been a lot of bashing of those instruments — some of it unjustified — but the fact is that an awful lot of institutions are still using off-the-shelf instruments without examining them. The point is not that we shouldn't use them at all, but that we don't generally examine the assumptions, the learning theory, and the psychometric assumptions that underlie these instruments when we decide to use them. Most of these tests, for instance, are built on the assumption that there are abstract *things* out there called "critical thinking" or "knowledge of American history" or something of that sort, things that are additive and that really exist in an ideal form that these tests somehow approximate. We don't often examine that kind of underlying assumption.

MORAN: Another problem with traditional approaches is that they produce an excessive focus on instruments and methods. The danger here is twofold. One is that the substantive meaning that the instruments are supposed to represent often gets lost in discussions of the technical adequacy of the instruments — How well is the question phrased? What kinds of results do you get from a factor analysis? and so forth — without a very clear understanding of what it is that we seek in the first place. The second is the danger of reification that I mentioned earlier. Like that Vonnegut character, the researcher becomes shielded from any understanding of the outcomes except as they are revealed through an analytical process.

What do I mean by that? Simply this: The whole process becomes rule-bound, cookbook-like. The researcher does whatever is necessary in the analysis to produce an outcome. He or she is unaware of what that outcome will be, because it's rule-bound and often highly mathematical. The coding that has gone into it quantifies the data, and that obscures a basic fact. The fact is that the researcher has asked questions that rest on his or her values. But those values are obscured by the end point.

So it's helpful to make the links between values and choices more explicit, in order to examine the connections between purpose and method more carefully. If we make that shift, it becomes possible for us to give as much weight to the process of self-discovery as to the results of the process. That's a real advantage. It helps us see assessment as connected to, rather than separated from, the daily experience of teaching and learning.

ASTIN: A hidden issue here is whether the assessment activity in question is intended to improve learning directly or indirectly. By "directly" I mean that the assessment results constitute immediate feedback to the learner and/or instructor that can directly enhance the very learning process. "Indirect" feedback improves learning by "enlightening" the educator or policy maker about which educational policies and practices are more or less effective. Direct feedback is often more "clinical," personal, and designed with the individual in mind. Indirect feedback may be more "researchy," more of a synthesis, and contain more aggregate information.

External agencies are interested almost exclusively in indirect uses of assessment. While it's theoretically possible to use the same assessment information for both purposes, we ordinarily employ quite different methodologies in the two situations. Aggregate information from standardized tests and even the individual scores, for example, are usually used indirectly, while final exams in a course are almost always used for direct feedback to the individual student. The point here is that this distinction in the *intent* of the assessor is often unclear. Unless assessors are very clear about this fundamental distinction from the beginning, they run a significant risk of choosing the wrong techniques and using them inappropriately.

MENTKOWSKI: Right, Sandy. The purposes we have for assessment make a big difference. If purposes aren't clear, we can create problems for ourselves. We can also

get into an endless debate about which level in the institution should be emphasized in designing assessment: individual, course, curriculum, department, institution, state, nation? Should we focus on the role of assessment in the *individual* learner's development, or in *system* improvements? Those with responsibility for a particular level of education (the student, the classroom instructor, curriculum designers, the academic dean, the president and board, the state or federal department of education) can easily be in conflict with one another about where assessment should be centered. So it helps to be clear about purposes.

It also helps to be clear about our values. For example, I hold the value that, irrespective of the level at which we are involved in assessment, all of us ultimately should be working to improve learning at the level of the individual student. That's why I agree with you, Sandy, that we need to build assessment systems that are integral to learning and that the students experience directly as part of their learning. We also need to be very clear about the kind of feedback we are generating, who will be able to *use* the information, for what purpose, and at what level. Our conversation about what students should learn and how well they are learning it should become an integral part of public discussion about educational goals and standards, as well as an expected part of instructor-student interaction about learning progress.

Another problem that arises, one related to what you said a bit ago about the importance of connections, Tom, is an excessive focus on assessment as a researcher's or an administrator's activity rather than as an educational activity. For example, instruments and methods are things. An assessment office is a place. You can use these instruments and these methods in this place over here, and then too easily forget about them!

Instead, we need to put more emphasis on assessment as an integral part of the *educational* process and ask ourselves what, exactly, is driving assessment. Is assessment just an administrative priority that we can stop worrying about because someone else is taking care of it? Or is assessment being driven by the educational process, along with the assumptions, values, and questions of the faculty? An excessive focus on instruments and methods, a view of assessment as an administrative activity, can distract us from the real purpose of assessment.

MORAN: Which is, of course, to improve the teaching/learning process. All too often, assessment appears quite distinct from that. Research suggests quite convincingly that knowledge cannot be separated from the context in which it is acquired or applied. Yet that's often what we attempt to do with traditional assessment strategies.

EWELL: Why do we do that? Because we're forced to, day in and day out, by the methods that we use to produce information and by the fact that we are for the most part using the information for accountability. The reason why many individuals are involved in assessment, although we would like to be involved for other reasons, is because someone is telling us to measure outcomes for public purposes. Marcia,

you mention the problem of assessment being an "add-on" activity; that problem becomes more acute when the stakes are high and you stand to win or lose a lot on the basis of some highly abstracted notion of gain.

The irony of all this, as I learn repeatedly in my conversations with public officials, is that they *want* us to be improvement-oriented. In our dealings with them, we tend automatically to think what they want is a *number,* while what they really are after is for us to look at ourselves more critically and to make the necessary improvements. In fact, our initial bureaucratic response to the assessment requirements placed on us from the outside often reminds me of the reactions of our students — "Is it going to be on the test?" "How long does it have to be?" — when we should see the point of the enterprise as continuous growth, challenge, and action.

ASTIN: Again, I think you are all talking about this distinction between direct use of results for "feedback" and indirect use for "enlightenment." Faculty involvement and "ownership" is almost a given when assessment is used as direct feedback right in the learning situation to pinpoint particular strengths and weaknesses, but not when administrators or outside agencies impose assessments for "enlightenment" purposes, that is, for making more general, public statements about effectiveness. It's also important to realize that the methodological requirements should be much more stringent when assessment is used for enlightenment purposes, because their basic intent is to establish causal connections between particular educational practices (environments) and particular educational outcomes. Even with the most sophisticated I-E-O design, however, faculty resistance is almost inevitable when assessments for "enlightenment" are sponsored by outside agencies.

MENTKOWSKI: That resistance is understandable when an excessive focus on accountability to "outsiders" leads us back to some of the old assumptions from testing practice. For example, the purpose of testing is often to make some kind of selection — who's acceptable, who's not — without substantive feedback. The purpose of assessment, in contrast, should always be to provide direct, substantive feedback, whether to the student or department, the faculty, institution, state board, or whatever. All too often, though, there's little or no feedback on any level.

So instead of generating results that can actually be used for improvement, assessment just generates information for some external purpose, and then assessment becomes indistinguishable from testing. Faculty resist, appropriately so. The fundamental question — *Why* are we doing this? — becomes obscured and we default to old modes of thinking and acting.

MORAN: That also leads, on many campuses, to the feeling that assessment is nothing more than a burden imposed by some other constituency. It hasn't yet become a central part of the culture of inquiry on the campus that you described earlier, Marcia. Historically, we *have* had assessment in the form of grading at the micro-level, where instructors do care about individual students, evaluate their work, and

talk to them about how they're doing; but we haven't had it at the macro-level — that is, assessment of how well the institution *as a whole* is achieving its purposes, by using student performance data on large numbers of students.

It's part of the culture of higher education to engage in assessment of individual students. It's not part of that culture to engage in assessment of large groups of students as a means of determining the value and efficacy of the curriculum or program, distinct from the individual student's ability to master the material.

MENTKOWSKI: And that, in turn, can lead to an excessive focus on off-the-shelf designs. Not just off-the-shelf instruments, but off-the-shelf designs. The pre-post design is a case in point. It's familiar, but when we use it, we can easily be tempted to neglect the program and process variables that Tom and Sandy have been talking about. Rather than focusing on what is happening in the educational process between the pre- and the post-test, and how that's connected to learning, there's a focus on the design itself.

EWELL: One of the things that you've all been talking about is that we generally see assessment activities as discrete pieces that ought to add up to something. But we tend to create these pieces without connecting them to one another. Different instruments, different approaches, and so on. One of the major ironies in all this is that the assessment movement was originally inspired, in part, by a search for coherence. Look at the central role assessment plays in reports such as *Involvement in Learning* or *Integrity in the College Curriculum* — reports that brought assessment to national attention. The whole movement in many respects was *about* a search for coherence in learning, but the measurement devices that we use tend to fragment things instead.

MORAN: Related to that problem is another one illustrated by my Vonnegut character: All too often in assessment, we fail to consider a sufficiently broad context. Assessment tends to focus primarily on cognitive outcomes of higher education; it doesn't examine the ways in which cognitive outcomes interact with other kinds of processes such as ethical development. Nor does assessment consider the way the individual interacts with the institution in a broader social context. You could say that the way we go about assessment decontextualizes student learning.

ASTIN: You've put your finger on a critical issue, Tom: our neglect of "affective" outcomes. In one sense the goals of a liberal education are at least as "affective" as they are "cognitive." Otherwise why do outcomes like "citizenship," "character," and "social responsibility" occupy such a central place in our catalogs and mission statements? There are at least two problems here. First, we shy away from including affective outcomes because we think they are difficult to assess: Where is a standardized multiple-choice test for citizenship? Yet the fact is that it is very easy and inexpensive to assess a wide range of affective outcomes through questionnaires and inventories. The assessments are rough and inexact, but they produce extremely useful information. Second, our curriculum doesn't really reflect our

supposed commitment to affective outcomes like citizenship. Here's a case where faculty brainstorming about general-education outcomes might well lead to curricular revision and reform. The important thing is to make sure that the brainstorming takes place in the context of a comprehensive and coherent sense of the institution's educational purpose and mission.

What I am really talking about here is our *values*. When you get right down to it, the things about ourselves that we try to measure or assess are a reflection of our values: These are the things we think are important. By implication, then, the things we *don't* assess are the things we value less! Does this mean that we don't really value citizenship, given that it's nowhere to be found either in our curriculum or in our assessment instruments? Should we delete citizenship from our catalogs and mission statements? These are the kinds of questions that get raised when we take a value-based approach to assessment.

MENTKOWSKI: So we need to pay attention to coherence and context, in light of our educational values. Higher education needs to expand its understanding of curricular coherence from just describing an institutional mission in the course catalog, or spelling out course sequences in the major, to articulating the abilities that cross the curriculum and connect general education to the major fields. Peter mentioned earlier that we tend to go at assessment piecemeal. I think we understand why that happens: When we're starting up assessment at the institutional level, we often have just these broad mission statements to go by. So we get something going over here and something else over there . . . it becomes a scatter plot design, where you can't draw relationships between any of these pieces or link them to a set of explicit assumptions about how students learn and how you want them to "turn out." Instead, what we want is a connect-the-dots picture, where if you work carefully, you actually can find the elephant.

To be able to connect the dots, we need to think about our goals, yes, but also our purposes, values, and underlying philosophy. Are we interested only in new applications of old assumptions — or are we genuinely interested in transforming higher education through assessment? If we're committed to transformation, that means reexamining our conceptual frameworks, thinking through the problems and limitations they pose, and finding new frameworks that offer solutions. But we're not just looking for new frameworks; we're also looking for the *connections* that *link* our new educational commitments to our assessment principles and practices, from the classroom and the campus to the statehouse. That's what this struggle for conceptual frameworks is all about.

III. ALTERNATIVES

MENTKOWSKI: Well then, what *are* some sources for new, more appropriate conceptual frameworks for assessment, for theoretical assumptions, that we could develop some consensus around? Speaking from my own perspective, when I connect

assessment to a conceptual base for myself, naturally I'm connecting it to the educational assumptions and frameworks that have evolved over twenty years at the institution where I work, Alverno College. I hope it's very clear to people, when they look at our assessment processes, that the coherence between our educational frameworks and assessment designs is deliberate: There is a dynamic interplay between our notions about ability-based, experiential learning and our development of performance-based assessment, for example.

Lately, though, I've also begun to identify some of the educational assumptions that are beginning to drive higher education assessment more broadly. I can see three intertwined, but distinct aspects: First, expanding the outcomes of college to include not only what students *know* but also what they are able to *do* has led to development of alternative assessments, including performance and portfolio assessment. Second, expanding learning to include active collaboration with others, and more reflective and self-sustained learning, has led to assessment of projects produced by groups of students and to more attention to self-assessment. Third, expanding educational goals to include personal growth has led to assessment of broad developmental patterns over time and to in-depth interviewing of students and alumnae. These educational assumptions and their practical implications for assessment find support in the literature, too. Let me elaborate.

First, as I mentioned a moment ago, there are the educational and psychological assumptions underlying *outcomes.* Many of us in our institutions are making outcomes more explicit by describing processes that link knowledge and performance. We're defining these processes as complex abilities that we expect our students to master by the end of general education or the major. We're calling those abilities "critical thinking" or "ethical decision making," and when we define such processes, we're clearly thinking of them as multidimensional rather than as simple, unitary constructs.

But those abilities are more than multidimensional; they're holistic. They include qualities of the person. They include not just knowledge or skills but attitudes, behaviors, even dispositions. We're beginning to understand that something like critical thinking has cognitive, affective, social, even kinesthetic dimensions. Moreover, we define those abilities as transferrable and we expect them to last a lifetime, to transfer across multiple aspects of work, family, and civic life long after college.

Clearly, the definitions of such outcomes are expanding, thanks to faculty teaching experience, but also thanks to psychological research. Take Lawrence Kohlberg and Carol Gilligan, for example. They expanded the definitions of moral reasoning in *Essays on Moral Development: The Philosophy of Moral Development* and *In a Different Voice,* respectively. Similarly, Muriel Bebeau and James Rest have expanded our definitions of moral sensitivity — see Rest's *Moral Development: Advances in Research and Theory.* All this work demonstrates that something like moral development or ethical decision making is multidimensional.

Similarly, our definitions of critical thinking are expanded by new definitions of intelligence. For example, there's George Klemp and David McClelland's definition of competence in *Practical Intelligence: Nature and Origins of Intelligence in the Workaday World,* edited by Robert Sternberg. Howard Gardner's multiple *Frames of Mind* and Sternberg's triarchic theories of intelligence in *Beyond IQ* further expand our picture of human potential. Each of these psychologists describes complexities that run counter to the notion of critical thinking as some kind of unitary construct. And yet the majority of psychometric measures and procedures currently available still treat these multidimensional constructs as unitary.

I want to digress for a moment and talk about the implications of this complexity. One implication for assessment design is that the more complex the ability, the more the student will benefit from faculty making what they mean by "abilities" explicit and public prior to assessment, for example, by delineating performance criteria. The student is also helped when these criteria describe sequential levels so that a student understands what distinguishes beginning from advanced performance.

Still another implication is that faculty assessment designers have come to rethink the question: What is "good" evidence? They are no longer satisfied with student selection of predetermined test items, or even short answers; they want essays, speeches, interviews, the critical incident, the journal, the lab experiment, even the recital or sculpture — active performances that are open, dynamic, and sustained. Expert judgment is needed to assess such performance, which implies that faculty within and across departments work together to figure out how to interpret this new kind of evidence. These discussions lead faculty to rethink how they want their students to turn out.

Many faculty feel they're just beginning to define abilities such as critical thinking in a particular major, and faculty are unwilling to limit their definitions just because measurement strategies are underdeveloped. Faculty are also finding that defining complex abilities and exploring how students learn them draws faculty deeper into thinking about how to assess those abilities. The upshot is that faculty resist — quite rightly — assessments imposed from the outside that tap only unidimensional abilities, or call for restricted student responses, just so outsiders can have aggregate information for accountability purposes. Thus, a shift in how to define complex outcomes — which is where we started — leads to all kinds of implications for assessment design and supports the longer-term investment that alternative forms of assessment can take.

I've just been talking about the assumptions underlying outcomes; now, as my second example, let's look at a set of assumptions about how we understand *learning*: Learning is a complex process that happens *experientially,* linking knowledge to action. Pat Hutchings and Allen Wutzdorff's *Knowing and Doing: Learning Through Experience* contains examples of that; so does David Kolb's theory in *Experiential*

Learning: Experience as the Source of Learning and Development. A number of colleges such as Lesley, Evergreen, and Fairhaven are experimenting with collaborative learning, trying to discover what really happens to the student in interacting with others, and how to assess learning outcomes.

Some psychologists have defined adult learning as socially constructed, as described in *Women's Ways of Knowing*, by Mary Field Belenky, Blythe McVicker Clinchy, Nancy Rule Goldberger, and Jill Mattuck Tarule, and, of course, in work by William Perry. In-depth, longitudinal interviews with Alverno alumnae led to this insight: Alverno graduates transfer college-learned abilities by using the learning skills they developed as students. Students and alumnae attributed the development of these skills, which included self-assessment and using feedback to improve, to their involvement in assessment at Alverno. When such learning is self-sustained and internalized, graduates can use it as a bridge to adjust and develop their abilities in many settings: in graduate school, as you might predict, but also at work, in the family, in civic roles, and even for changing careers. Here, then, is an implication for designing assessments that is emerging from both our educational practice and research: Students *can* transfer their learning and abilities when they are assessed in situations that are similar to those that they will perform in later. This means carefully reviewing the kinds of activities students are asked to perform when they are assessed, and developing a more advanced picture of the complex abilities they are likely to need after college.

My third example centers on *student development*. Here I see a well-articulated assumption coming out of faculty practice and the literature: The aims of education should include *developing the whole person*, as well as learning a discipline or profession. Curricula at a number of institutions build in a developmental approach: Millsaps College, Miami University of Ohio, Alverno, Empire State, and DePaul's School for New Learning are just a few. Ernest Pascarella and Patrick Terenzini, in *How College Affects Students*, review twenty years of college outcomes research and make the case that we should think about *change* during college as *development*. Those two words are not necessarily synonymous. In this view, change is a kind of cognitive development in which *structural* shifts in thinking and learning occur. Some faculty I know are talking about structural shifts in affective development, as well. When change or development is defined as nonlinear, as qualitative shifts in deep structure rather than surface structure, that has profound implications for how we will go about measuring or demonstrating change, particularly when what we really are after is insight into the *patterns* of change that occur *as* the student is learning.

Good teachers and college student personnel have taught us that this kind of deep-structure development is a central aim of education — indeed, it's the theme of Arthur Chickering's *The Modern American College*. Many faculty are thinking about how to assess for that. For example, in the major you would look not only for growth in disciplinary understanding but also for personal growth, and for how the two interact. That leads us back to your earlier observations, Peter, about how

we measure change. I think we need to expect *change* in complex abilities, as students benefit from instruction, rather than consistency in some underlying "trait" over time. That's a shift in one of our basic measurement assumptions, a shift that has enormous implications for assessment.

EWELL: Let's dissect this last concept a bit. How are we actually able to detect or model either change or development? What does a path of development look like? Paths of development often may have regressions. They may loop back to earlier conditions. They may take us back to revisit old assumptions and see old things in new ways. They may result, ironically enough, in *no detectable change* whatsoever on a central-tendency measure. You may completely change the way in which you think about a concept and yet score the same on a given examination — and that's one of the major limitations in the way we've traditionally looked at these things.

ASTIN: Our most recent research on student development suggests that some of the most interesting change in a cohort of students is in the *variability* rather than in the mean. This appears to be happening in such diverse domains as quantitative skills and political beliefs, where students become more *diverse* over time.

MENTKOWSKI: What you've both just said suggests that when we think about new assessment technologies, we really need to look at those that will pick up on the intraindividual, or within-the-individual, patterns of change. We have to assume change rather than consistency; expect nonlinear as well as additive change; and look for group-level patterns, without obscuring individual differences. To return to earlier themes, what patterns in the development of complex, higher-order abilities are multidimensional and transferable beyond college? Right now, we don't have too many ways of pulling those patterns apart.

EWELL: But we can be practical about it. In some cases it means taking an existing instrument and looking at the items as well as the underlying scales. I once looked at James Rest's "Defining Issues Test" that way. As you know, it's been used a lot in trying to measure moral judgment and has also been criticized for too heavy a reliance on Kohlberg's model of moral development. Rest's items are embedded in a situation or story, though, so it's possible to re-cluster the items to explore other aspects of moral reasoning. When you look at how the individual items are behaving, you may arrive at a different pattern or configuration from what the underlying scales suggest. It's fascinating.

ASTIN: This is an extremely important issue when you're dealing with standardized tests. The "feedback" from a percentile or standard score has very little pedagogical value, but the results on individual items can be of tremendous value. You can see not only *which* items students are having problems with but which wrong alternatives ("distractors") they are choosing. I wish the testing companies would routinely feed back response distributions on individual items, because "total" scores don't really tell you much and can even be misleading.

MENTKOWSKI: That kind of analysis is just as important when analyzing the criteria from performance assessments, where criteria are a set of behavioral descriptions of what you expect students to be able to demonstrate with respect to a certain ability like critical thinking. You look at the criteria, and then you look at a particular student's performance, and you may find that someone who is doing a unique project — a poem, say — meets those criteria in quite unique ways. You try to get a feel for the particular pattern or mix of criteria that one student demonstrates versus the pattern of another student. There may be many different pathways. And then you also ask yourself what you're not seeing, and should be seeing, given how you define the ability you're looking for.

For example, in one class the question we started with was: "Are our minority students doing as well in our biology classes as our other students?" The instructor was very concerned because some of the students were not passing the assessments. When we started to unpack that, we discovered that there were many different ways in which students solved problems in biology, but the professor was teaching — and assessing — for only a few of those. So the instructor articulated a larger number of strategies, and that has really opened up what she has been able to do with that particular class.

EWELL: Let me play devil's advocate. Is it still important that students, *all* of them, be able to solve the problem?

MENTKOWSKI: Very definitely, particularly if they're going into nursing. I think, Peter, if you were in the hospital, you would very much like those Alverno nurses to be able to solve those biological problems.

EWELL: So we need assessment instruments that can do both jobs — that can at the same time tell us *where* students are with respect to what we want them to know and *how* they are learning.

MENTKOWSKI: Exactly.

MORAN: There's another problem, though: We're not only imposing a single model of assessment on students; we're imposing it on faculty, as well. I think one of the reasons why assessment has failed to have the impact on faculty cultures that we hoped it would is that we're not allowing other ways of knowing to surface in the assessment process. There's a hegemony of traditional psychometric theory; other ways of knowing that are characteristic of other disciplines — for example, the humanities — are not seen as relevant or valid in generating assessment data. Here again, we need to change our thinking: Instead of relying on the behavioral sciences for all our models, we need to promote interdisciplinary approaches to assessment design.

MENTKOWSKI: That's why it's so important to focus on learning, and how we think students learn, when we're designing assessment. Learning seems to be one of the

"outcomes" that we can look at across the disciplines, as Bill Perry did. At Alverno, we've done deep-structure interviews with students from many disciplines to see how they construct learning, how they construct meaning. As Lee Shulman has argued, the way students construct meaning about learning may be quite different from one discipline to another, and yet there are probably some common themes, as well. We've found that it's quite possible to track different ways in which students construct meaning about learning and about their college experience across an entire curriculum.

EWELL: How do we continue that conversation about different ways of learning? Probably the most important question we have to face is that there are some elements that are going to be common across the learning process, and others that are going to be highly discipline-specific. Shall we start with the disciplines, where faculty actually live? Or with the few things that faculty may agree should be critical outcomes of the college experience? Or somewhere else?

MORAN: We need to start, I think, with a critique of traditional empiricism, or more precisely, positivism, which underlies psychometric theory and has driven so much of assessment practice to this point. We need to articulate some assumptions that have practical application in designing assessment, and that open us up to a wider range of disciplinary perspectives.

It seems to me there are three central tenets in positivism, or traditional empiricism. First of all, it embodies a view of knowledge as truth: as an external, extant, verifiable reality waiting to be discovered. That has relevance to the way we approach assessment, which I'll get to in a moment. Second, this traditional empirical world view holds that reality is governed by a set of laws that are predictable, mechanistic, and deterministic. These laws dictate certain outcomes. This, too, has relevance for the way we look at our students and the way they learn and develop. Finally, the way we know or discover this reality or truth is through sensory evidence — observation — indeed, this is the methodological foundation of empiricism. Of course, there *are* other means of knowing truth and discovering knowledge, but they are not included in positivism.

What are the weaknesses of positivism? It has little tolerance for the unique or the particular, little use for context; such phenomena are treated as error variance, not worthy of study. Individuals are viewed as subject to certain laws, their behavior prescribed by external forces. When we look at our students through this methodological lens, we are looking for those conditions that control their behavior; we are *not* looking at what students bring to the setting, nor are we looking at the student as a self-willed actor.

Traditional empiricism ignores the highly complex interactions that develop as a result of the interplay between social actors and the historical and social context in which they live. Empiricism, as we've said several times, has a tendency to decontextualize. By failing to take into account historical and social context, tra-

ditional empiricism downplays continuity and its influence in shaping human beliefs or behavior. Hence, empiricism bedevils our efforts to understand human growth — the ultimate goal of educational assessment.

In traditional empiricism, instrumental or technical reasoning dominates over other forms of inquiry and judgment. Because empiricism downplays historical and social context, there's too little consideration of "community" as a source of knowledge, values, or behavior. The instrumental logic of empiricism leads us to focus on the *means* by which we know something, rather than focusing on the substantive questions we are attempting to resolve.

Nowadays, of course, traditional empiricism, or positivism, is under attack. Where are these attacks coming from, and what are the alternatives? I alluded earlier to the theoretical developments in the social sciences that took place in the late 1960s and early 1970s. One milestone was the publication of *The Structure of Scientific Revolutions,* by Thomas Kuhn; another was Berger and Luckmann's *The Social Construction of Reality.*

These thinkers challenged the notion of knowledge as objective reality. They emphasized that our knowledge of the world is shared and reproduced in human consciousness through dialogue — a process referred to as "intersubjectivity." They argued that we don't so much "discover" the world as construct and enact it. Their work took place in the context of twentieth-century revolutions in physics and art that led us to see the world as relativistic, contingent, and in nonlinear patterns.

That revolution has now taken hold, it seems to me, in a wide range of disciplines. Again, three things characterize this revolution. One, the notion of fixed truths: Objectivity has been challenged. Two, the notion of simple cause-and-effect models, or a mechanistic view of the world, has been challenged. And three, the view that individual behavior is reducible to a deterministic pattern; the notion that there are deterministic rules that explain the world and that only need to be discovered — that, too, has been called into question.

How has this played out in some of the traditional disciplines? There's Peter Novick's book, *That Noble Dream,* in which he takes on the objectivity question in history. In philosophy, we have the work of Richard Rorty; his books *Philosophy and the Mirror of Nature* and, more recently, *Contingency, Irony, and Solidarity* present a powerful view of the contingent nature of knowledge and the central importance of community (rather than empirical method) in validating understanding. The play of multiple perspectives can also be seen in such books as *Literary Theories in Praxis,* edited by Shirley Staton.

Two lines of work deserve a more detailed review, however. One is the work of sociologist/philosopher Jürgen Habermas. Habermas acknowledges that empiricism can be undeniably useful for ascertaining certain qualities about reality, but he argues that its usefulness is limited. He does *not* simply reject a notion of objec-

tive reality and embrace a notion of subjective reality. That's a dichotomy we fall into too easily when we talk about assessment, as Marcia noted earlier. We need a view of assessment that unifies those elements.

Incidentally, aligning qualitative research with subjectivity and quantitative research with objectivity is misguided. Qualitative and quantitative techniques are both related to empiricism in that they rely, first, on observation and, second, on "technical" standards for the gathering, presenting, and justifying of evidence on which conclusions are based. The difference between qualitative and quantitative approaches lies in the *way* the two approaches determine the questions of fundamental importance, and in the way the meaning of conclusions is negotiated during the research process.

In assessment, as in human life generally, questions about meaning and values rest on broader forms of reasoning than the techniques of empiricism. But when empirical techniques or instrumental reasoning dominates human discourse, the ways in which we arrive at judgments about values and meanings are seen as inferior to the ways we arrive at empirically derived "facts." The consequence is that values and meanings are not considered "real knowledge"; there's no place for them in (supposedly) value-neutral systems of knowledge.

This in part explains a pervasive phenomenon in educational measurement; I'm talking about our tendency — traditionally, at least — to measure what's easy rather than the more-complex outcomes that are harder to assess. Once we start down this path, the consequences are multiple: We come to value what we measure, instead of measuring what we value, and we send that skewed message to other constituencies. The trivial can too easily displace what is truly important. Moreover, we never do develop the language to talk about many of those things we genuinely value, nor do we develop the strategies to provide evidence of them. In a sense, we fail to legitimate those more complex educational outcomes, and so they disappear from our educational landscape.

What's the remedy? Habermas suggests that we integrate the *means* of knowing — that is, our assessment strategies — with the substantive *purposes* of knowing, which derive from our educational values; and he urges us to focus on community as a source of knowledge.

A second challenge to empiricism, in addition to Habermas, comes from anthropology and is represented most notably by the work of Clifford Geertz in *The Interpretation of Cultures*. The fundamental concern of the anthropologist is to understand action *from the point of view of the social actor*. That's very different from traditional educational research.

As I said earlier, I'm *not* talking here about qualitative versus quantitative research. The distinction I'm making here is between positivist research strategies and ethnography. Positivist research is unilateral: The researcher makes assumptions

a priori, and then seeks evidence to confirm or disconfirm those assumptions. Ethnographic research, in contrast, attempts to understand the world as the social actors in it understand it. Ethnography strives to understand how those actors make sense of their world. That forces the researcher into a position of reciprocity: The researcher and the social actors *negotiate* the meaning of that world.

Those, I think, are some of the intellectual sources that will help to provide new conceptual frameworks for assessment. We've seen practice run ahead of theory; now I think we need theory to catch up with practice.

MENTKOWSKI: I agree that assessment practice has run ahead of those theories. For example, those of us doing assessment certainly don't think of, say, a general-education committee or an assessment practitioner as somehow standing outside of the institution, like the old-fashioned empiricist, or even a newfangled ethnographer, who travelled to distant cultures but remained an "outsider" who later wrote a book and returned to his or her faculty position. The assessment practitioner *is* a member of the community, and likely to remain so.

EWELL: Good point. One pioneering anthropologist — either Bronislaw Malinowski or Alfred Radcliffe-Brown, I think — said about his colleagues in the field that the really good ones never came back, that "going native" was the ultimate in successful understanding. As I was listening to Tom's summary, particularly the references to Habermas, it struck me that our entire discussion reveals the field of assessment itself to be sitting somewhere in the middle on the Perry scheme of intellectual growth. We've moved away from a notion of revealed truth, of right and wrong answers, of linear testing methodologies as the only way to go. And now we are in a multiplicity stage: We see that diversity is legitimate. Right now, every method may seem as good as every other method. There are few, if any, rules of conduct. Anything goes.

MORAN: We haven't gone far enough yet. We haven't linked this diversity to conceptual frameworks, to rationales, to evidence.

ASTIN: Once again, the problem is that we can easily confuse the narrow issues of instrumentation and measurement procedure and technique with the larger questions: Why are we doing this? What do we hope to learn? How might we use the resulting information? How will we make sense out of it? And how can we get the larger academic community to take an interest in it, ascertain meaning from it, and use it to improve student learning and development? My own strong feeling is that we shouldn't even consider the questions of instrumentation and methodology until we've at least *tried* to answer such questions.

EWELL: That's right. How can we, in Perry's language, move to *commitment?* What will we commit to? What are the standards of evidence we should be talking about? What constitutes good evidence? What is our theory of evidence? We need to address these questions systematically. For illustration, let me lay on the table a set

of questions that I think everybody engaging in assessment ought to be able to think through. Most of the time we don't, in fact, engage in this exercise, but it can be very revealing. We should go through such a list every time we pick up an instrument, whether we take it off the shelf or design it ourselves from first principles.

I had occasion to go through an exercise like this in a piece of writing I did recently. For better or worse, I came up with three basic questions that we simply *have* to ask ourselves about anything that we do in the realm of evidence gathering or measurement. They are (1) What is the construct? (2) What is the context? (3) What are you going to *use* the results for?

Embedded in that first question, the one about the construct, are some further issues. Just what abilities *are* we trying to measure? What are the attributes? What do we really want evidence about? Are we talking, for instance, about a property of an individual or are we also talking about a *process* of learning? Are we talking about an activity that an individual *engages in*, as well as a quality the individual *has*? If it's a property that we're talking about — the way people often talk about "critical thinking," for example — is it a generalized attribute, something that everybody should have "some of" along the same metric, or are we also trying to describe how an ability is learned in a particular context? How you conceptualize an assessment process makes a huge difference in the kinds of evidence that make sense and the kinds of measurements that are going to work.

By and large, up to now, we've tended to model such things as though they were generalized abilities. For example, we assumed there was such a *thing* as critical thinking out there that you had some of, then more of, and we conceived of every individual as "having it" in more or less the same way. If we conceived of it that way, central-tendency measures made sense. Some abilities may, in fact, be like that — for example, most basic skills and arguably some metacognitive skills as well. The point here is that *how* you model the ability is an active *choice* that you have to make — and then you have to choose measures or methods that make sense in relation to the model. You can't let the measure make the choice for you.

Interestingly enough, there are some alternative paradigms out there for certain abilities. In talking with assessors of writing, for example, Edward White has concluded that "inter-rater reliability" may need to be rethought in some judgment situations. If you talk to three expert judges of a given piece of writing, they may not think of the differences among their ratings as "error." Instead, they think of the differences in the scores that they award as reflecting different interpretations of the same phenomenon. And that is a very different way to put things together than a psychometric scale. Nonetheless, White would agree that assessors must reach some kind of consensus when there are explicit consequences for the student, such as advancement or graduation. Again, we've got to constantly be thinking about what the assessment we're doing is *for*.

If you are talking about a process rather than a property — that is, if you conceive of what you are looking for as a path of development — then how do you model that? Is it linear? More of the same? An additive model, where you begin with some of it, and then get more and more of it? In that case, you can look at an individual as though he or she were being filled up like a bucket. There are some knowledge areas that are like that, where you might want to look for change by taking repeated observations. But there may be others where the changes are *structural,* where the changes lie in the ways particular elements of knowledge are related to one another, where the model of "filling up a bucket" doesn't work at all.

When there has been structural change, for instance, you may get the same overall score on a given measure, even though within the items that comprise it, a change in configuration has occurred. This can happen in lots of different kinds of settings. One of the most interesting that I've encountered was at a college that wanted to examine religious faith among its students and developed an instrument to look at the tenets of belief within a particular religious system. In fact, and not surprisingly, they got a "value-subtracted" result. Over the course of their education at this institution, students became less and less willing to say that every element of the church's creed was what they believed.

But more interestingly, the instrument went on to inquire about the *basis* of their belief. Did they believe what they did because of what their parents told them? Was it because they had analytically considered a set of ethical or theological questions? Was it because of scripture itself? How did they come to these conclusions? What the college was looking for was not so much a change in whether students believed in particular things, but rather change in the *structure* of the belief itself. They wanted to understand why students constructed things the way they did.

So that's the first question, the one about construct. The second question is: What is the context? Can complex abilities, if they do exist in this abstract form, be abstracted from the specific situations that they arise in? What kinds of techniques make sense in translating from one context to another?

Here there are at least two kinds of contexts that we need to think about. One of them is situational: the actual ground on which the thing is taking place. If, as with Marcia's nursing students, we're talking about a practice context, we'd better not abstract much further than that. But we at least would like the assessment to be constructed to cross the different practice settings that the student is likely to encounter.

The other kind of context to keep in mind is that of individual difference. We have to ask ourselves: What effect is this question or approach going to have across genders, ethnic groups, and — most importantly, perhaps — across learning styles and across different ways of seeing the world. You *have* to address that question, and make sure that you're not abstracting too far from where the abilities you want

to look at are actually manifested.

The third question is the one that really fascinates me, though: What are you going to *use* the information for? It's important to recognize that even at a place like ETS, which we tend to think of as pretty traditional, Samuel Messick argues that the whole notion of validity is tied to the way results are used — that validity depends as much on what you are going to *do* with the information as it does on the psychometric properties of the instrument that produced it.

That means thinking through some questions very concretely: How great is the risk that we might be wrong? What are the consequences if we're wrong? Who pays the price of being wrong? Those are all very useful questions to raise at the outset of an assessment program, because they force you to think through what you are going to do with the information once you've got it.

All three of these things — construct, context, and use — imply rethinking what constitutes good evidence and developing a specific *theory* of evidence. Every assessment practitioner ought to have such a theory, not just a tool-kit of measures.

MENTKOWSKI: I think, too, there's a major press right now to reconsider that question — What constitutes good evidence? — because of our orientation toward diversity and inclusion of multicultural dimensions in the curriculum, in assessment, and so on. Can we really get good evidence of an ability without considering the context in which the student has developed the ability over time, which might mean her cultural background? Can we get good evidence without understanding the affective dimensions of the student's performance? Can we get good evidence without considering the context in which the student will need to demonstrate the ability later, after college?

Let's go back to that biology class. Students from different backgrounds will have different affective responses when they demonstrate problem solving in a context, depending on the modes of problem solving that they grew up with. It all becomes very complex. When an ability like problem solving is multidimensional, that means we have to assess in quite different ways than we used to. What are the alternatives? How do we take into consideration the student's cultural background, and also the various cultural backgrounds of the persons she will be dealing with after college? I find the implications for assessment of these shifts in assumptions fascinating. What about the rest of you?

EWELL: Let me address your question indirectly. The major issue for me, as I've said, is thinking through any given assessment. That's my underlying message here. There's no single recipe that is going to work for everybody. You have to ask that set of questions I mentioned earlier: What's the construct? What's the context? What are the uses? The answers to those questions may be very different from campus to campus, or from department to department. You may end up with a

standardized test off the shelf, depending on the purposes that you are trying to accomplish. But you *have* to think through those choices from first principles: What is your underlying theory? Who is the audience? What are the results going to be used for? What is the risk of going wrong? and so on.

All that brings us back again to the question of evidence itself. "Evidence" is an interesting word. Evidence by its very nature implies cross-individuality, that is, being able to talk from one person to another. It implies persuasion. It implies being able to bring somebody into a community of judgment. That's a bridge to Tom's notion of intersubjectivity and the question of how we create a community of judgment. "Evidence" implies much more than just measurement.

MORAN: Yes, the notion of a community of judgment is absolutely central here. Your example of Messick's concern about validity — that the *use* to which the instrument will be put, not just its internal psychometric properties, is a determinate of validity — that's extraordinarily significant. It represents a fundamental inversion. Not long ago, a judgment about whether something was an adequate assessment device rested on technical properties alone, irrespective of the use to which it was put or the context in which student learning occurred.

This change requires dialogue, with the reciprocity that implies, between the assessment designers and all those who will use the information before validity can be ascertained. Dialogue as part of the research process and the importance therefore of having some sort of community in which that dialogue can take place are, it seems to me, the central changes now occurring in the way we look at research.

EWELL: It's also interesting to note that the specific debate that Messick was involved in when he reformulated the notion of "validity" grew out of a tremendous angst on the part of the measurement community regarding the misuse of testing results — in particular, the serious misapplication of standardized testing in elementary and secondary school systems. This was a real danger, and the test designers in their ivory tower realized that making a "good" test on the basis of psychometric principles alone just wasn't good enough.

So one of the things that we have to talk about is that ultimately the community of judgment includes people *outside* the academy who will use our results for different ends. We need to include them in the dialogue, as well.

MORAN: That's right, but this angst that you refer to on the part of the measurement community represents another issue, too. In *Fourth Generation Evaluation*, Egon Guba and Yvonna Lincoln discuss the phenomenon of interactivity in the research process; a similar concept, reflexivity, is further elaborated by Frederick Steier in *Research and Reflexivity*. That is, the researcher's own values and intentions shape the outcomes of research during the course of inquiry, and so they are also a legitimate source of investigation in the research process. And that, too, is

very different from the empirical research tradition, with its belief in value neutrality. That neutrality was always an illusion, of course; the researcher *did* have an influence on the outcomes, it was simply ignored or overlooked. Even the way in which the researcher asks a question prefigures the response or outcome.

MENTKOWSKI: All the more important, then, to carefully construct the community of judgment. When it includes students, faculty, departments, and the institution as a whole, the community of judgment, as I see it, contains within it both the contextual dimension, where values and consequences interact, *and* the desire to make comparisons against criteria that capture more than just one's own expertise and understanding. These criteria may be drawn from within this community at first, and then gradually expanded to include sources outside it, other groups and constituencies. For example, we include assessors from the business and professional community in our student assessment process. When we evaluate a major field, we may look to the disciplines or professions for broader criteria. As we examine how our curriculum contributes to lifelong learning, we may draw from descriptions of human potential across the lifespan or take readings from the settings in which our graduates perform. Regardless of how a community of judgment constructs itself, it can engage in continuous, public discussions about what the criteria and comparisons should be. As those become public, the fear about hidden educational goals or standards disappears. Then a student, department, or institution more or less knows at the start what the community of judgment expects, and it can construct performance in a way that will enable comparisons with the criteria that have been publicly discussed.

EWELL: Once again, we see that the community of judgment includes people *outside* the academy. It's essential to include them. And one of the things that we need to be aware of here is that their view of us and what they expect of us is shifting. All through the 1960s and 1970s, public higher education was viewed by public officials as a kind of *public utility* — funded and justified primarily on the basis of the benefits that it provided to individuals in the form of increased income, social mobility, quality of life, and so on. Now, more insistently, what we do in higher education is justified as a public *investment in the future* that will pay off in terms of economic growth, work force development, and functional citizenship. One consequence of this shift is that we're increasingly being asked, through assessment, to demonstrate return on investment more explicitly. But another consequence, directly related to our point about the community of judgment, is that determining the particular abilities and attributes that we must teach toward in the colleges and universities of the future is no longer exclusively *our* concern. Complex abilities such as critical thinking, effective communication, and problem solving — all mentioned in the National Education Goals — are part of an important and widening public conversation about what college graduates should know and be able to do. This is not a conversation that we as assessors can stand outside of, because its outcome will profoundly affect both what we are asked to assess and how the results will be used.

ASTIN: In thinking about this interface between the assessment practitioner and the larger community, we have to keep in mind what the current concept of "assessment" really *is* in that community. At this point in time, I'm not sure many members of that larger community would be able to follow this discussion, because *their* experience and expectation is that assessment means GPAs, final exams, SATs, ACTs, GREs, and other norm-referenced assessments. Their notion of it is that you give a one-shot test to see what people have "learned." They believe that "what is being tested is what has been learned." They think assessment should always be "cognitive," because "affective" outcomes are too "value-laden" and therefore none of the business of academia. Policy makers and the general public would not buy the argument that the "effectiveness" of an undergraduate program could be judged in terms of something like the Perry scheme; I'm not even sure many academics would buy it.

The point is that if we want to enlist the involvement and support of the larger academic and public communities, we have to look at this as an *educational* process, where we start where they are and *lead* them gradually into a more in-depth discussion of issues like longitudinal assessment, criterion-reference measurement, the role of values, affective outcomes, portfolios, performance assessment, ways of knowing, contextual assessment issues, and the like. We have to think more about how to get the discussion from there to here, how to communicate emerging conceptual frameworks, and how to get educators to think more about how assessment can be linked to these frameworks.

IV. NEXT STEPS

MENTKOWSKI: Given everything that's been said, where do we go from this conversation? I think we've demonstrated that reviewing assessment practice is a springboard for reexamining our educational mission and goals; the defining question there is "What do we want for students?" Now, as an assessment community, we are beginning to deal with "why" and "how" questions, linking them back to our central purposes. Why design assessment this way for what reasons? Why ask for this kind of evidence, and for whose use? How do we implement this kind of assessment process for what benefits? Who is learning from participating in assessment? If we reflect on our assessment practices along these lines, we may find our way to some more explicit guidelines, some "must-be-theres." Gradually, our conceptual frameworks may emerge more clearly. Asking "How is this different from what we did before?" leads to the uncovering of new elements and principles that distinguish what we hope to do from our earlier conventional practices. Asking "Why do we care?" generates the link back to our educational assumptions and values.

It seems to me that several activities, engaged in by all of us in the assessment community, might move us along. Here are a few; they could come in any sequence: We might generate some common guidelines for how to design and do

assessment more effectively. Individually, we could work at inferring the educational assumptions and assessment principles that seem to undergird how we do assessment in our own contexts, then look for links among our educational reasons, assessment design guidelines, actual practices, and apparent benefits. Informed by such an exercise, groups of us might sit down together and look for reasons why our principles and practices are similar or different.

This group activity could help to start the task of identifying the criteria that characterize "good" assessment in the many effective examples that we hear about from the AAHE Assessment Forum and in other settings. Finally, we could all participate in making explicit how assessment designs rest on those educational values and assumptions we are committed to, those that we are working toward, and those that we think are more central to where we hope to be in the future.

The goal of these activities is to communicate the interactive nature of practice and theory more clearly; to establish more explicit links among our educational values and commitments, and activities; and to make the connections among our educational assumptions and assessment principles more public.

As an assessment community, we may want at some point to design more systematic ways to carry on these activities. For now, analyzing our practices for insights about our educational assumptions and assessment principles is clearly intellectually stimulating, and it brings us into a multiplicity of interdisciplinary and intercampus discussions. In the interim, the real test of whether assessment rests on sound conceptual frameworks resides more in our daily practice, and less in ideal descriptions. Students, after all, experience how we do assessment, not necessarily what we intend. This conversation is an invitation for our readers' voices to join ours. ■

About AAHE

The American Association for Higher Education (AAHE) is a national organization of individuals dedicated to improving the quality of American higher education. Its members include faculty and administrators as well as educational policy makers. Through AAHE's conferences, publications, and special projects — such as the AAHE Assessment Forum — members can acquire both the "big picture" and the practical tools to increase their own effectiveness.

The benefits of membership in AAHE include subscriptions to *Change* magazine and the *AAHE Bulletin;* discounts on conference registration; discounts on selected non-AAHE subscriptions *(The Journal of Higher Education* and *ASHE-ERIC Higher Education Reports);* and more.

Use the coupon below to join.

AMERICAN ASSOCIATION FOR HIGHER EDUCATION

AAHE members receive free the *AAHE Bulletin* (ten issues/year) and *Change* magazine (six issues/year); significant discounts on conference registration; special rates on selected non-AAHE subscriptions; Hertz car rental discounts; and more. To join, complete this form and send it to AAHE, One Dupont Circle, Suite 360, Washington, DC 20036-1110.

Membership *(choose one)*
Regular: ☐ 1 yr, $75 ☐ 2 yrs, $145 ☐ 3 yrs, $215
Retired: ☐ 1 yr, $45 Student: ☐ 1 yr, $45
(For all categories, add $8/year for membership outside the U.S.)

Caucuses *(you must be an AAHE member; choose same number of years as new membership above)*

Amer. Indian/Alaska Native:	☐ 1 yr, $10	☐ 2 yrs, $20	☐ 3 yrs, $30
Asian/Pacific American:	☐ 1 yr, $15	☐ 2 yrs, $30	☐ 3 yrs, $45
Black:	☐ 1 yr, $15	☐ 2 yrs, $30	☐ 3 yrs, $45
Hispanic:	☐ 1 yr, $25	☐ 2 yrs, $50	☐ 3 yrs, $75
Lesbian/Gay:	☐ 1 yr, $10	☐ 2 yrs, $20	☐ 3 yrs, $30

Name (Dr./Mr./Ms.) _____ ☐ M/☐ F

Position _____

Institution/Organization _____

Address (☐ home/☐ work) _____

_____

City _____ St _____ Zip _____

☐ Bill me ☐ Check enclosed *(payment in U.S. funds only)*

BEST COPY AVAILABLE

About the AAHE Assessment Forum

A special project of the American Association for Higher Education and funded by FIPSE, the AAHE Assessment Forum was established in 1987 to help coordinate and advance thoughtful assessment. The Forum offers a variety of services and resources, including the annual AAHE Conference on Assessment, commissioned papers and conference presentations, national directories, an in-house assessment resource library, assessment resources via ERIC, assistance with consulting needs, and more.

For more information about AAHE or its Assessment Forum, contact AAHE, One Dupont Circle, Suite 360, Washington, DC 20036-1110; ph. (202) 293-6440; fax (202) 293-0073.

Barbara D. Wright
Director, AAHE Assessment Forum

Elizabeth A. Francis
Project Assistant, AAHE Assessment Forum

FORUM RESOURCES

Reprise 1991 **$14.00**
Reprints of two outstanding papers treating assessment's history, implementation, gains, and shortcomings, with examples of campus practice and extensive bibliography.
- "To Capture the Ineffable: New Forms of Assessment in Higher Education" by Peter T. Ewell (from *Review of Research in Education*)
- "Watching Assessment — Questions, Stories, Prospects" by Pat Hutchings and Ted Marchese (from *Change*)

Catching Theory Up With Practice: Conceptual Frameworks for Assessment (1991) **$10.00**
By Marcia Mentkowski, Alexander W. Astin, Peter T. Ewell, E. Thomas Moran, and K. Patricia Cross. A conversation about how assessment practice connects with changes in the ways we think about epistemology, student learning, measurement, and evaluation.

Using Assessment to Strengthen General Education (1991) **$10.00**
By Pat Hutchings, Ted Marchese, and Barbara Wright. How assessment's questions and approaches can support this central component of undergraduate education.

Assessment 1990: Understanding the Implications **$12.00**
From the 5th AAHE Conference on Assessment in Higher Education
- "Streams of Thought About Assessment" by K. Patricia Cross
- "The Truth May Make You Free, But the Test May Keep You Imprisoned: Toward Assessment Worthy of the Liberal Arts" by Grant Wiggins
- "Assessment and the Way We Work" by Pat Hutchings

Assessment 1990: Accreditation & Renewal **$10.00**
From the 5th AAHE Conference on Assessment in Higher Education
- "Assessment and Accreditation: A Shotgun Marriage?" by Ralph A. Wolff
- "Assessment as a Tool for Institutional Renewal and Reform" by Alexander W. Astin

Time Will Tell: Portfolio-Assisted Assessment of General Education (1990) **$9.00**
By Aubrey Forrest and a study group that included fifteen experts in the field. A comprehensive guide to the implementation and use of student portfolios to assess general-education outcomes at individual and program levels.

Assessment Programs and Projects: A Directory **$ 12.00**
Edited by Jacqueline Paskow, updated by Elizabeth A. Francis (1987, 1990). Concise descriptions of thirty assessment projects implemented on campuses across the country—including purposes, key features, strategies and instruments, and impact.

Behind Outcomes: Contexts and Questions for Assessment (1989) **$ 5.00**
By Pat Hutchings. This paper sets forth nine areas of inquiry for assessment that gets "behind outcomes," with appropriate methods for addressing each area and resources for further work.

Three Presentations—1989 **$10.00**
From the 4th National Conference on Assessment in Higher Education
- L. Lee Knefelkamp, "Assessment as Transformation"
- Peter T. Ewell, "Hearts and Minds: Some Reflections on the Ideologies of Assessment"
- Rexford Brown, "You Can't Get There From Here"

Three Presentations—1988 **$10.00**
From the 3rd National Conference on Assessment in Higher Education
- Alexander Astin, "Assessment and Human Values: Confessions of a Reformed Number Cruncher"
- Linda Darling-Hammond, "Assessment and Incentives: The Medium Is the Message"
- Robert H. McCabe, "The Assessment Movement: What Next? Who Cares?"

***Three Presentations—1987** **$ 5.00**
From the 2nd National Conference on Assessment in Higher Education
- Lee S. Shulman, "Assessing Content and Process: Challenges for the New Assessments"
- Virginia B. Smith, "In the Eye of the Beholder: Perspectives on Quality"
- Donald M. Stewart, "The Ethics of Assessment"

Resource Packet II: Six Papers **$25.00**
- "Acting Out State-Mandated Assessment: Evidence From Five States," C. Boyer and P. Ewell
- "Assessing Student Learning in Light of How Students Learn," J. Novak and D. Ridley
- "Faculty Voices on Assessment: Expanding the Conversation," P. Hutchings and E. Reuben
- "Feedback in the Classroom: Making Assessment Matter," K. Patricia Cross
- "Standardized Tests and the Purposes of Assessment," J. Heffernan
- "An Update on Assessment," (AAHE *Bulletin*, December, 1987), P. Hutchings and T. Marchese

***Resource Packet I: Five Papers** **$8.00**
- "Assessment, Accountability, and Improvement: Managing the Contradiction," P. Ewell
- "Assessment and Outcomes Measurement: A View From the States," C. Boyer, P. Ewell, J. Finney and J. Mingle
- "The External Examiner Approach to Assessment," B. Fong
- "Six Stories: Implementing Successful Assessment," P. Hutchings
- "Thinking About Assessment: Perspectives for Presidents and Chief Academic Officers," E. El-Khawas and J. Rossmann

*These publications are available only as photocopies.

BEST COPY AVAILABLE

ORDER FORM

Name _____

Position _____

Institution _____

Address _____

City _____ State _____ Zip _____

FORUM RESOURCES

No. of Copies	Title	Total
_____	Reprise 1991 ($14)	_____
_____	Catching Theory Up With Practice ($10)	_____
_____	Using Assessment to Strengthen General Education ($10)	_____
_____	Assessment 1990: Understanding the Implications ($12)	_____
_____	Assessment 1990: Accreditation & Renewal ($10)	_____
_____	Time Will Tell ($9)	_____
_____	Assessment Directory, rev 1990 ($12)	_____
_____	Behind Outcomes ($5)	_____
_____	Three Presentations - 1989 ($10)	_____
_____	Three Presentations - 1988 ($10)	_____
_____	Three Presentations - 1987 ($5)	_____
_____	Resource Packet II: Six Papers ($25 set)	_____
_____	Resource Packet I: Five Papers ($8 set)	_____
	Delivery (see below)	_____
	Resources Subtotal	_____

Orders under $50 must be prepaid; orders of $50 and over must be accompanied by payment or purchase order. Make checks payable to **AAHE Assessment Forum**. Book-rate postage is included in prices above. Allow four weeks for delivery. For first-class, UPS, or foreign delivery, add $2 for 1 copy ordered; $2.50 for 2-4, and $3.50 for 5 or more copies. No refunds or exchanges.

JOIN AAHE

Regular membership ☐ $75 1 yr ☐ $145 2 yrs ☐ $215 3 yrs
Special membership (1 yr only) ☐ full-time student $45 ☐ retired $45
☐ Please send AAHE membership information

For all categories, add $8 per year for membership outside the U.S.

Membership Subtotal _____

TOTAL ENCLOSED _____

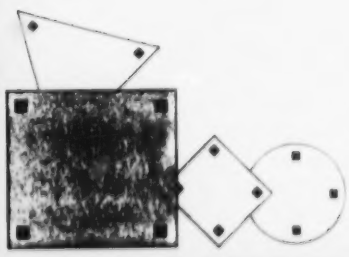
Please add the following names/addresses to the AAHE Assessment Forum mailing list:

_____

_____

_____

Mail completed form and payment to:
AAHE Assessment Forum, One Dupont Circle
Suite 360, Washington, DC 20036



AAHE
AMERICAN ASSOCIATION FOR HIGHER EDUCATION

American Association for Higher Education
One Dupont Circle, Suite 360
Washington, DC 20036-1110
Ph. (202) 293-6440 Fax (202) 293-0073